ON A GENERAL ASYMPTOTIC INDEPENDENCE RESULT IN STATISTICS

By Bo Bergman

Saab-Scania AB

In a recent paper Barlow and Proschan noted that similar independence results appeared both in life table analysis and in fixed interval analysis. In this note we present a general asymptotic independence result of which these results are special cases. Also some further applications are given. In essence, our method may be described as follows: Assume that we are interested in g quantities q_1, \dots, q_g , each constructed from a sequence of independent random variables, and that these sequences are conditionally independent given their random lengths. Then by the completion of each sequence with independent random variables we obtain g independent sequences. Under rather general assumptions we are able to deduce asymptotic properties of the original sequence from the corresponding properties of the completed sequences. In particular, we are often able to prove the asymptotic independence of the quantities q_1, \dots, q_g .

1. Introduction. In a recent paper, Barlow and Proschan (1976) noted that similar asymptotic independence results appeared both in life table analysis (cf. Chiang, 1968, Breslow and Crowley, 1974, and Barlow and Proschan, 1977) and in fixed interval analysis (cf. Sethuraman, 1961, 1963a, b). The methods of proof were quite different. In this note we shall show that the above results are special cases of a more general asymptotic independence result.

In life table analysis the positive real line is partitioned into intervals $I_{\ell} = (a_{\ell-1}, a_{\ell}]$, $\ell = 1, \dots, g$, with $0 = a_0 < a_1 < \dots < a_g \le \infty$. Given n independent observations from the life distribution F(F(0-) = 0), let F_n be the corresponding empirical measure. Usually the parameters of interest in life table analysis are the conditional probabilities of death, $q_{\ell} = F[I_{\ell}]/\bar{F}(a_{\ell-1})$, estimated by $\hat{q}_{\ell} = F_n[I_{\ell}]/\bar{F}_n(a_{\ell-1})$, $\ell = 1, \dots, g$; here $\bar{F} = 1 - F$ and $F[I_{\ell}] = F(a_{\ell}) - F(a_{\ell-1})$. Chiang (1968) shows that these estimators, suitably normalized, are asymptotically independent and normally distributed. (Actually, he also shows that, conditionally on $F_n[I_{\ell}] > 0$ and $F_n[I_{\ell}] > 0$, \hat{q}_{ℓ} and \hat{q}_{ℓ} are uncorrelated for finite samples.) Breslow and Crowley (1974) extend the asymptotic result to the case of random censorship.

The asymptotic independence of certain reciprocals of sample means has been noted by Barlow and Proschan (1977). In particular, they show that the normed interval failure rate estimators,

(1.1)
$$\sqrt{n} \left[F_n[I_\ell] \middle/ \int_{a_{\ell,l}}^{a_\ell} \bar{F}_n(s) \ ds - F[I_\ell] \middle/ \int_{a_{\ell,l}}^{a_\ell} \bar{F}(s) \ ds \right], \qquad \ell = 1, \dots, g,$$

are asymptotically independent and normally distributed.

In a series of papers on fixed interval analysis Sethuraman (1961, 1963a, b) shows the asymptotic independence of similar conditional sample means based on empirical measures. Let (Y, X) be a random vector taking values in $(R^k \times \mathcal{X})$, where R^k is the Euclidean space of k dimensions and \mathcal{X} is the set $\{x_1, \dots, x_g\}$. Furthermore, let (Y_i, X_i) , $i = 1, \dots, n$, be independent random elements (r.e.'s) distributed as (Y, X) and assume that the covariance matrix for Y is finite and that $P(X = x_i) > 0$, $\ell = 1, \dots, g$. Let F_n denote the

Received August 1980; revised August 1981.

Key words and phrases. Asymptotic independence, fixed interval analysis, life table analysis, reliability.

AMS 1980 subject classifications. Primary 60F05; secondary 62N05.

empirical measure on *X*. Sethuraman (1963b) shows that

(1.2)
$$\sqrt{n} \left[\frac{1}{n} \sum_{\{i \mid X_i = x_\ell\}} \mathbf{Y}_i / F_n[x_\ell] - E(\mathbf{Y} \mid X = x_\ell) \right], \qquad \ell = 1, \dots, g,$$

are asymptotically independent and normally distributed.

The aim of this paper is to present a general asymptotic independence result applicable both to life table analysis and to fixed interval analysis. In essence, our method may be described as follows: Assume that we are interested in g quantities q_1, \dots, q_g , each constructed from a sequence of independent r.v.'s, and that these sequences are conditionally independent given their (random) lengths. Then by the completion of each sequence with independent r.v.'s we obtain g independent sequences. Under rather general assumptions we are able to deduce asymptotic properties of the original sequences from the corresponding properties of the completed sequences. In particular, we are often able to prove the asymptotic independence of the quantities q_1, \dots, q_g .

In Section 2 we give the general result and in Section 3 we prove the asymptotic independence result in fixed interval analysis due to Sethuraman (1963a). We also indicate an extension of that result. In Section 4 we study life table analysis proving the results of Breslow and Crowley (1974) and Barlow and Proschan (1977); the latter result is extended to the case of random censorship. Finally, in Section 5 we use our general method to deduce an asymptotic independence result in the theory of semi-Markov processes. This result is of importance in e.g. reliability theory.

In the following we shall assume that a probability space, (Ω, \mathscr{F}, P) is given and that it is rich enough to make all studied random elements measurable with respect to the Borel Field $(B.F.)\mathscr{F}$.

2. The general result. Let \mathscr{F}_n , $n \ge 1$, be an increasing sequence of B.F.'s included in \mathscr{F} , and (η_n, ξ_n) , $n \ge 1$, a sequence of independent, identically distributed (i.i.d.) random elements (r.e.'s) adapted to \mathscr{F}_n such that η_n takes values in some measurable product space $\mathscr{Y} = \times \mathscr{E}_{=1} \mathscr{Y}_{\ell}$, and ξ_n is a random binary g-dimensional vector such that $P(\xi_{\ell n} = 1) > 0$, $\ell = 1, \dots, g$.

Assumption A0. We assume that $\eta_{1n}, \dots, \eta_{gn}$, the components of η_n , are conditionally independent given ξ_n .

For each $\ell = 1, \dots, g$ and $i \ge 1$ let $h_{\ell i}$ be a measurable function on the product space \mathcal{Y}_{ℓ}^{i} taking values in some metric space \mathcal{H} endowed with the topology induced by the metric:

$$h_{\ell i}: \mathscr{Y}^{\iota}_{\ell} \rightsquigarrow \mathscr{H}, \qquad \ell = 1, \cdots, g.$$

Also, let $\nu_{\ell n}$, $\ell=1, \cdots, g, n \geq 1$, be integer valued functions measurable with respect to \mathscr{F}_n such that $\nu_{\ell n} \leq n$ and let $\mu_{\ell n} = \sum_{i=1}^{\nu} \xi_{\ell i}$. Here and in the following, we supress the references to ℓ and n in $\nu = \nu_{\ell n}$ and $\mu = \mu_{\ell n}$ whenever the risk of misinterpretation is negligible. In most applications we have $\nu \equiv n$. Now, for each ℓ ($\ell=1, \cdots, g$), let $\bar{\eta}_{\ell n}, n \geq 1$, be the thinned sequence constructed from the sequence $\eta_{\ell n}, n \geq 1$, by taking away those components $\eta_{\ell j}$ for which $\xi_{\ell j} = 0$. We shall be interested in the asymptotic properties of the r.e.'s.

(2.1)
$$\bar{u}_{\ell n} = h_{\ell \mu}(\bar{\eta}_{\ell 1}, \cdots, \bar{\eta}_{\ell \mu}), \qquad \ell = 1, \cdots, g;$$

for an interpretation of these quantities in e.g. fixed interval analysis, see the beginning of Section 3.

In order to obtain our main result in this section we shall introduce g auxiliary independent sequences of independent r.e.'s $\zeta_{\ell l}$, $i \geq 1$, $\ell = 1, \dots, g$, taking values in \mathscr{Y}_{ℓ} , being independent of $\bigcup \mathscr{F}_n$, and each with distribution equal to the conditional distribution of $\eta_{\ell 1}$ given $\xi_{\ell 1} = 1$. We may now for each $n \geq 1$ and $\ell(\ell = 1, \dots, g)$ construct a new sequence of r.e.'s:

$$\zeta_{\ell i}^{(n)} = \begin{cases} \bar{\eta}_{\ell i}, & i = 1, \dots, \mu \\ & , \qquad n \geq 1. \end{cases}$$

$$\zeta_{\ell i}, \quad i = \mu + 1, \dots$$

By conditioning on ξ_j , $j=1, \dots, n$, it is immediately seen that for each n the g sequences $\zeta_{\ell i}^{(n)}$, $i \geq 1, \ell = 1, \dots, g$, are independent with i.i.d. components distributed as $\zeta_{\ell 1}$. Obviously, we may write

$$\bar{u}_{\ell n} = h_{\ell \mu}(\zeta_{\ell 1}^{(n)}, \cdots, \zeta_{\ell \mu}^{(n)}), \qquad \ell = 1, \cdots, g.$$

We are now ready to make the following basic assumptions.

Assumption A1. Assume that for each ℓ , $\ell = 1, \dots, g$, we have, as $n \to \infty$ (with the usual abuse of notations)

$$u_{\ell n} = h_{\ell n}(\zeta_{\ell 1}, \cdots, \zeta_{\ell n}) \rightarrow_D G_{\ell};$$

here \rightarrow_D denotes weak convergence and G_ℓ , $\ell=1, \cdots, g$, are probability distributions on \mathcal{H} .

Assumption A2. Assume that for each $\ell = 1, \dots, g$, a fixed sequence of integers $m = m_{\ell n}, n \ge 1$, exists such that, as $n \to \infty$, we have $m \to \infty$ and

$$\|\bar{u}_{\ell n}-h_{\ell m}(\zeta_{\ell 1}^{(n)},\cdots,\zeta_{\ell m}^{(n)})\|\rightarrow_P 0.$$

From the independence of the sequences $\zeta_{\ell i}^{(n)}$, $i \geq 1$, $\ell = 1, \dots, g$, and from Assumption A1 we have that $h_{\ell m}(\zeta_{\ell 1}^{(n)}, \dots, \zeta_{\ell m}^{(n)})$, $\ell = 1, \dots, g$, are asymptotically independent with asymptotic marginal distributions G_1, \dots, G_g . Now, from Theorem 4.1 in Billingsley (1968) and Assumption A2 it follows that $\bar{u}_{1n}, \dots, \bar{u}_{gn}$ are asymptotically independent with asymptotic marginal distributions G_1, \dots, G_g . That is the main result of this section. In the following we shall exploit this simple but powerful result in a number of applications.

REMARK. It is obvious that at the price of heavier notations we could have formulated the above results by using triangular arrays.

3. Fixed interval analysis. In order to prove the asymptotic independence result in fixed interval analysis due to Sethuraman (1963a), cf. Section 1, we shall give suitable interpretations of the quantities in Section 2. Let

$$\eta_{\ell n} = \begin{cases} \mathbf{Y}_n - E(\mathbf{Y}_n \mid X = x_{\ell}) & \text{if} \quad X_n = x_{\ell} \\ 0 & \text{otherwise,} \end{cases}$$

and

$$\xi_{\ell n} = \begin{cases} 1 & \text{if } X_n = x_{\ell}, \\ 0 & \text{otherwise.} \end{cases}$$

Then obviously the components η_n are conditionally independent given ξ_n ; thus Assumption A0 is fulfilled. Furthermore, let $\nu \equiv n$ and let the functions $h_{\ell n}$ be defined by

$$h_{\ell n}(y_1 \cdots, y_n) = \sqrt{n}(\sum_{i=1}^n y_i/n).$$

Hence from the central limit theorem we know that Assumption A1 is fulfilled and that G_{ℓ} , $\ell=1,\cdots,g$, are normal distributions in R^k . Also Assumption A2 is fulfilled for $m=\lfloor nP(X=X_{\ell}) \rfloor$; here $\lfloor z \rfloor$ denotes the largest integer smaller than or equal to z. This follows easily by a use of the Kolmogorov inequality (see e.g. the proof of Theorem 7.3.2 in Chung, 1974). The assumptions of Section 2 are fulfilled and hence the asymptotic independence result follows from the observation that with the given interpretations, the quantities in (1.2) are equal to those in (2.1) multiplied by quantities converging to positive constants with probability one.

The methods originally used by Sethuraman (1961, 1963a) consist, in essence, in conditioning on the r.v.'s $t_{\ell} = \sqrt{n} \left(\mu_{\ell n} / n - P(X_1 = x_{\ell}) \right)$ and proving what is called UC* convergence (weak, uniform and continuous convergence) with respect to t_{ℓ} of the conditional random vector $(\bar{u}_{1n}, \dots, \bar{u}_{gn})$. Using a limit theorem for joint distributions (Sethuraman, 1961), Sethuraman then deduces the given result.

Closely related results may easily be obtained by our technique. Assume e.g. that the actual number of observations is random and that, based on the first i-1 observations and on X_i , one may decide not to observe Y_i . This will not cause any but trivial changes in the above results as long as the total number, $\mu_{\ell n}$, of observations contributing to the conditional sample mean, given $X = x_{\ell}$, fulfils $\mu_{\ell n}/n \to \theta_{\ell}$ as $n \to \infty$. The described situation may occur if e.g. observations are costly.

4. Life table analysis.

4.1. Preliminaries. Before applying the results of Section 2 to various life table situations, we shall give a basic result. Assume the model given in Section 2 and let $\mathscr{Y}_{\ell} = R^2$ and $\eta_{\ell n} = (\tau_{\ell n}, \delta_{\ell n})$. Later, we shall interpret $\tau_{\ell n}$ and $\delta_{\ell n}$ in life table situations as certain characteristics of the *n*th individual in the ℓ th time interval. We shall assume that the random vector $(\tau_{\ell n}, \delta_{\ell n})$ has finite second moments, and that ξ_n fulfils the requirements of Section 2.

Now, let $\zeta_{\ell i} = (t_{\ell i}, d_{\ell i})$ be defined as in Section 2, and let

$$u_{\ell n} = h_{\ell n}(\zeta_{\ell 1}, \dots, \zeta_{\ell n}) = \sqrt{n} \left(\sum_{i=1}^n d_{\ell i} / \sum_{i=1}^n t_{\ell i} - E(d_{\ell i}) / E(t_{\ell i}) \right).$$

Then

(4.1)
$$u_{\ell n} = \frac{\{E(t_{\ell 1})\sum_{i=1}^{n} (d_{\ell i} - E(d_{\ell 1})) - E(d_{\ell 1})\sum_{i=1}^{n} (t_{\ell i} - E(t_{\ell 1}))\}/\sqrt{n}}{E(t_{\ell 1})\sum_{i=1}^{n} t_{\ell i}/n},$$

and thus, it is easily seen that $u_{\ell n}$ is asymptotically normal with mean zero and standard deviation σ_{ℓ} , where

(4.2)
$$\sigma_{\ell}^{2} = \frac{E^{2}(t_{\ell 1})V(d_{\ell 1}) + E^{2}(d_{\ell 1})V(t_{\ell 1}) - 2E(d_{\ell 1})E(t_{\ell 1})\operatorname{cov}(t_{\ell 1}, d_{\ell 1})}{E^{4}(t_{\ell 1})}.$$

Hence Assumption A1 is fulfilled. We also have, if $\mu/n \to_P \theta_\ell > 0$ as $n \to \infty$, that

$$|u_{\ell[n\theta_{\ell}]} - u_{\ell\mu}| \to_P 0.$$

The proof of this is a simple application of the Kolmogorov inequality (cf. the proof of Theorem 7.3.2 in Chung, 1974). Hence also Assumption A2 is fulfilled and $\bar{u}_{\ell n}$, $\ell = 1$, ..., g, are asymptotically independent and normally distributed. We are now ready to prove asymptotic independence results in various life table situations.

4.2. The Breslow-Crowley result. Breslow and Crowley (1974) studies a model with random censorship, i.e. each random observation is bivariate, (Z_n, Δ_n) , where Z_n may be interpreted as

$$Z_n = \min(Z_n^0, X_n)$$

and

$$\Delta_n = egin{cases} 1 & ext{if} \quad Z_n = Z_n^0 \ 0 & ext{otherwise}; \end{cases}$$

here Z_n^0 has c.d.f. F, and X_n , the censoring variable, has c.d.f. H. The quantities of interest are

$$q_{\ell} = F[I_{\ell}]/\bar{F}(a_{\ell}), \qquad \ell = 1, \dots, g;$$

here I_{ℓ} and a_{ℓ} are defined as in Section 1, and for any c.d.f. G we use the notation $\bar{G} = 1 - G$. For more details on this model see Breslow and Crowley (1974). We shall interpret the variables in the last subsection in the following way:

$$\delta_{\ell i} = \begin{cases} 1 & \text{if } Z_i \in I_{\ell} \text{ and } \Delta_i = 1 \\ 0 & \text{otherwise,} \end{cases}$$

and

$$\tau_{\ell i} = \xi_{\ell i} - w_{\ell i}/2;$$

here

$$w_{\ell\,i} = egin{cases} 1 & ext{if} \quad Z_i \in I_\ell \quad ext{and} \quad \Delta_i = 0 \ 0 & ext{otherwise,} \end{cases}$$

and

$$\xi_{\ell\,\iota} = egin{cases} 1 & ext{if} & Z_i > a_{\ell-1} \ 0 & ext{otherwise}. \end{cases}$$

The traditional life table estimator of q_{ℓ} is given by

$$\hat{q}_{\ell} = \frac{\sum_{i=1}^{n} \delta_{\ell i}}{\sum_{i=1}^{n} \tau_{\ell i}}.$$

Hence from the last subsection we infer the asymptotic independence and normality of the quantities

$$\sqrt{n}(\hat{q}_{\ell}-q_{\ell}^*), \qquad \ell=1, \cdots, g;$$

here $q_{\ell}^* = E(\delta_{\ell 1})/E(\tau_{\ell 1})$. The asymptotic variance is easily calculated from (4.2).

4.3. The Barlow-Proschan result with random censorship. We shall generalize the Barlow-Proschan result to the case of random censorship, i.e. we shall use the same model as in the last sub-section, but we shall interpret $\tau_{\ell,t}$ as

$$\tau_{\ell i} = \int_{0}^{Z_{i}} 1_{I_{i}}(s) \ ds;$$

here $1_I(\cdot)$ is the indicator function of the set I. In life testing terminology $\tau_{\ell \iota}$ may be interpreted as the total time generated by the ith individual in the interval I_{ℓ} . A natural (naive) estimator of the interval failure rate $\bar{\lambda}_{\ell}$,

$$\tilde{\lambda}_{\ell} = F[I_{\ell}] / \int_{a_{\ell-1}}^{a_{\ell}} \bar{F}(s) ds,$$

is

$$\hat{\bar{\lambda}}_{\ell} = \frac{\sum_{i=1}^{n} \delta_{\ell i}}{\sum_{i=1}^{n} \tau_{\ell i}}.$$

From Subsection 4.1 we know that

$$\sqrt{n}(\hat{\bar{\lambda}}_{\ell}-E(\delta_{\ell 1})/E(\tau_{\ell 1})), \qquad \ell=1, \cdots, g,$$

are asymptotically independent and normally distributed, if the first and second moments of $(\delta_{\ell i}, \tau_{\ell i})$ exist and are finite. This is trivially the case for $\ell \leq g-1$. We assume that the requirement is fulfilled also for $\ell = g$. It is of some interest to investigate the consistency

of $\hat{\bar{\lambda}}$. From the definitions of $\delta_{\ell 1}$ and $\tau_{\ell 1}$ we obtain

$$E(\delta_{\ell-1}) = \int_{a_{\ell-1}}^{a_{\ell}} \bar{H}(s) \ dF(s)/(\bar{H}(a_{\ell-1})\bar{F}(a_{\ell-1})),$$

$$E\left(au_{\ell\,1}
ight) = \int_{a_{\ell-1}}^{a_{\ell}} ar{H}(s)ar{F}(s) \; ds/(ar{H}(a_{\ell-1})ar{F}(a_{\ell-1})).$$

Hence it is easily seen that $\hat{\bar{\lambda}}_{\ell}$ is a consistent estimator of $\bar{\lambda}_{\ell}$ if the quantity B,

$$B = \bar{\lambda}_{\ell} - \frac{E(\delta_{\ell 1})}{E(\tau_{\ell 1})} = \frac{\int_{a_{\ell - 1}}^{a_{\ell}} dF(s)}{\int_{a_{\ell - 1}}^{a_{\ell}} \bar{F}(s) \ ds} - \frac{\int_{a_{\ell - 1}}^{a_{\ell}} \bar{H}(s) \ dF(s)}{\int_{a_{\ell - 1}}^{a_{\ell}} \bar{H}(s) \bar{F}(s) \ ds},$$

is equal to zero. Since

$$B = \frac{\int_{a_{\prime-1}}^{a_{\prime}} \bar{H}(s)\bar{F}(s) \ ds \int_{a_{\prime-1}}^{a_{\prime}} dF(s) - \int_{a_{\prime-1}}^{a_{\prime}} \bar{F}(s) \ ds \int_{a_{\prime-1}}^{a_{\prime}} \bar{H}(s) \ dF(s)}{\int_{a_{\prime-1}}^{a_{\prime}} \bar{F}(s) \ ds \int_{a_{\prime-1}}^{a_{\prime}} \bar{H}(s)\bar{F}(s) \ ds},$$

we have B=0 if \bar{H} is constant in I_{ℓ} , i.e. if there is no censoring in I_{ℓ} , or if the failure rate function, $r(s)=f(s)/\bar{F}(s)$ (assuming the pdf f exists) is constant on I_{ℓ} . It is also easily seen that $\hat{\lambda}_{\ell}$ is biased if e.g. the failure rate function is strictly monotone. The case of no censoring was studied by Barlow and Proschan (1977) and the case of constant failure rate in each interval was studied by Crow and Shimi (1977).

5. Semi-Markov processes. We may apply our result in Section 2 also to the case when e.g. a semi-Markov process is observed during a long time period. We give the following simple formalization: Assume that the imbedded Markov chain takes a finite number of states, $S_1, \dots S_{\lambda}$, and that the holding time in state S_i has cdf G_i . Now, assume that we want to estimate the transition probabilities of the Markov chain and the means of the holding times after a long period of observation of the process. Assuming that the second moments of the holding times exist an almost straight-forward application of the results in Section 2 provides us with the asymptotic distribution of the vector of maximum likelihood estimators suitably normalized, as T, the time of observation, tends to infinity.

It might be worth noting that the alternating renewal process is a special case. Alternating renewal processes occur naturally in the study of availability measures in reliability theory, cf. e.g. Barlow and Proschan (1975).

Acknowledgement. I wish to thank professor R. E. Barlow, who drew my attention to asymptotic independence results. I also wish to thank Drs. J. Lanke and G. Lindgren for their kind interest in my research and a referee for some helpful comments.

REFERENCES

Barlow, R. E. and Proschan, F. (1975). Statistical Theory of Reliability and Life Testing. Holt, Rinehart and Winston, New York.

Barlow, R. E. and Proschan, F. (1976). Multivariate life table analysis. ORC 76-9, Operations Research Center, University of California, Berkeley. Barlow, R. E. and Proschan, F. (1977). Asymptotic theory of total time on test processes with applications to life testing, in *Multivariate Analysis IV* (ed. Krishnaiah) 227–237. North-Holland, Amsterdam.

BILLINGSLEY, P. (1968). Convergence of Probability Measures. Wiley, New York.

Breslow, N. and Crowley, J. (1974). A large sample study of the life table and product limit estimates under random censorship. *Ann. Statist.* 2 437-453.

CHIANG, C. L. (1968). Introduction to Stochastic Processes in Biostatistics. Wiley, New York.

CHUNG, K. L. (1974). A Course in Probability Theory. 2nd ed., Academic, New York.

Crow, L. H. and Shimi, I. N. (1977). Estimation of a multiparameter failure rate using a general replacement plan, in *The Theory and Applications of Reliability II* (ed. Tsokos and Shimi). Academic, New York.

RAO, C. R. (1973). Linear Statistical Inference and its Application. 2nd ed. Wiley, New York.

SETHURAMAN, J. (1961). Some limit theorems for joint distributions. Sankhyā, Ser. A 23 379-386.

Sethuraman, J. (1963a). Some limit distributions connected with fixed interval analysis. Sankhyā, Ser. A 25 395-398.

SETHURAMAN, J. (1963b). Fixed interval analysis and fractile analysis, in *Contribution to Statistics* (ed. C. R. Rao). Pergamon Press, Oxford and Statistical Publishing Society, Calcutta.

SAAB-SCANIA AB S-581 88 LINKÖPING SWEDEN